

Center for Basic Algorithms Research Copenhagen



2017 VILLUM Investigator

Mikkel Thorup

Information and Communication Technologies (ICT)

- a vehicle for societal progress



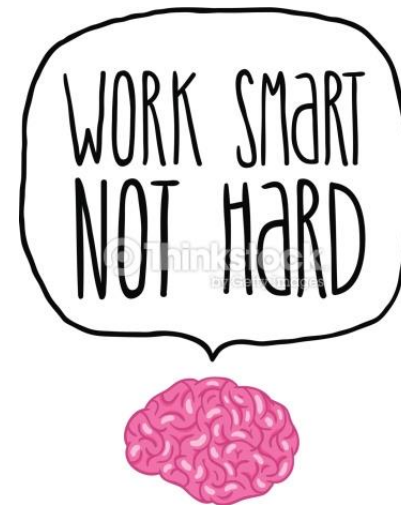
“Information and Communication Technologies underpin innovation and competitiveness across private and public sectors and enable scientific progress in all disciplines.”

“The ICT sector represents 4.8% of the European economy.”

“Investments in ICT account for 50% of all European productivity growth.”

Algorithms are key to keeping up progress

The amount of data grows much faster than computer speeds, so need for efficient algorithms to process data becomes more and more urgent.



Example: Vimeo

Main competitor of YouTube – 170 million users/month.
Serves about 1 billion requests for video clips per day.



The screenshot shows the Vimeo website interface. At the top, the Vimeo logo is on the left, followed by navigation links: 'Join' (in a green button), 'Log in', 'Host videos', 'Watch', and 'On Demand'. A search bar contains the text 'Search videos, people, and more' with a magnifying glass icon. To the right of the search bar is an 'Upload' button with an upward arrow icon.

The main video player displays a video titled 'THE FOURTH INDUSTRIAL REVOLUTION'. The video frame shows a dark, abstract background with glowing blue lines and a central vertical beam of light. A 'STAFF PICK' badge is visible in the top-left corner of the video frame. On the right side of the video frame, there are icons for heart, clock, list, and share. Below the video frame is a progress bar showing '11:34' and a play button icon. To the right of the progress bar are icons for HD, a monitor, and a full-screen icon.

Below the video player, the video title 'The Fourth Industrial Revolution' is displayed in large, bold, dark blue text. Underneath the title, it says 'from Marta Chierago BUSINESS 4 months ago | more'. To the left of the title is a circular profile picture of Marta Chierago, and to the right is a '+ Follow' button.

At the bottom left of the video player area, there are engagement metrics: a play button icon followed by '72.8K', a heart icon followed by '2,039', and a comment icon followed by '111'. Below these is a speech bubble icon followed by '36'. To the right of these metrics is a 'Share' button with a right-pointing arrow icon.

On the right side of the page, there is a section titled 'Search results for "algorithms"'. Below the title is a toggle switch for 'Autoplay next video' which is currently turned on. Below this is a search result card for the video 'The Fourth Industrial Revolution' from Marta Chierago, which includes a small thumbnail of the video.

Key technology: Consistent hashing

Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications

Ion Stoica*, Robert Morris, David Karger, M. Frans Kaashoek, Hari Balakrishnan†
MIT Laboratory for Computer Science
chord@lcs.mit.edu
<http://pdos.lcs.mit.edu/chord/>

Abstract

A fundamental problem that confronts peer-to-peer applications is

to d
pap
this
a k
imp
iter
key
sys
cha
per

and the state maintained by each node scaling logarithmically with the number of Chord nodes.

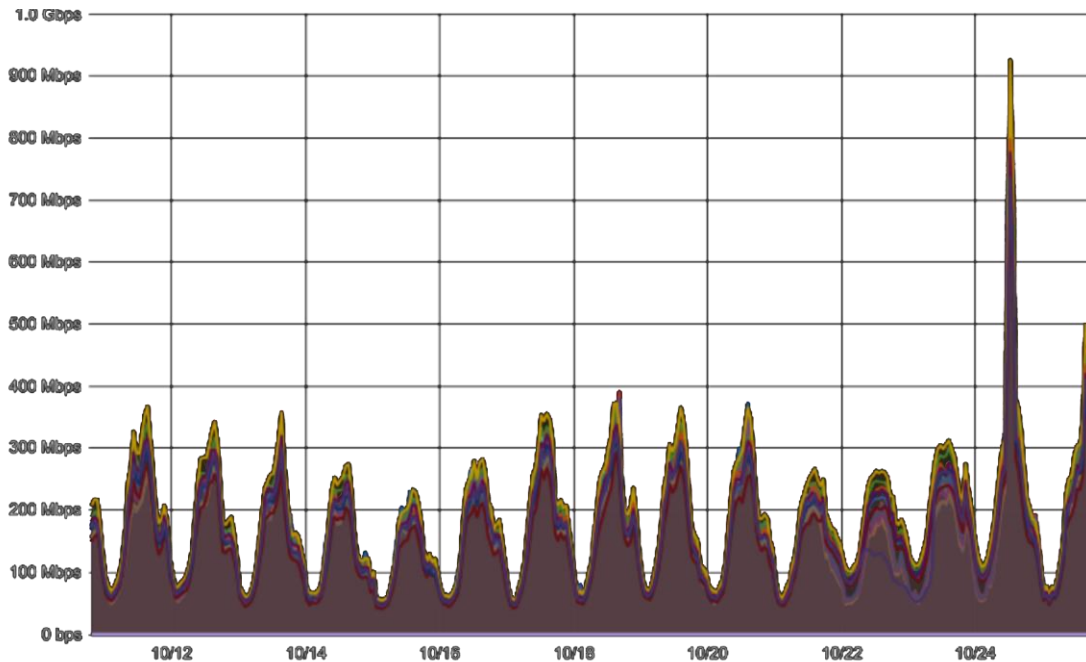
and involves relatively little movement of keys when nodes join and leave the system.

event results in no more than $O(\log^2 N)$ messages.

Three features that distinguish Chord from many other peer-to-peer lookup protocols are its simplicity, provable correctness, and

Title	1–20	Cited by	Year
Chord: A scalable peer-to-peer lookup service for internet applications		12552	2001
I Stoica, R Morris, D Karger, MF Kaashoek, H Balakrishnan ACM SIGCOMM Computer Communication Review 31 (4), 149-160			

Vimeo's bandwidth bottleneck



Issue: High bandwidth requirement...

From algorithm theory to industrial reality

Vimeo Engineering Blog

Follow



Improving load balancing with a new consistent-hashing algorithm

We run Vimeo's dynamic video packager, Skyfire, in the cloud, serving almost a billion DASH and HLS requests per day. That's a lot! We're very happy with the way that it performs, but scaling it up to today's traffic and beyond has been an interesting challenge. Today I'd like to talk about a new algorithmic development, *bounded-load consistent hashing*, and how it eliminates a bottleneck in our video delivery.



Cornell University
Library

arXiv.org > cs > arXiv:1608.01424v1

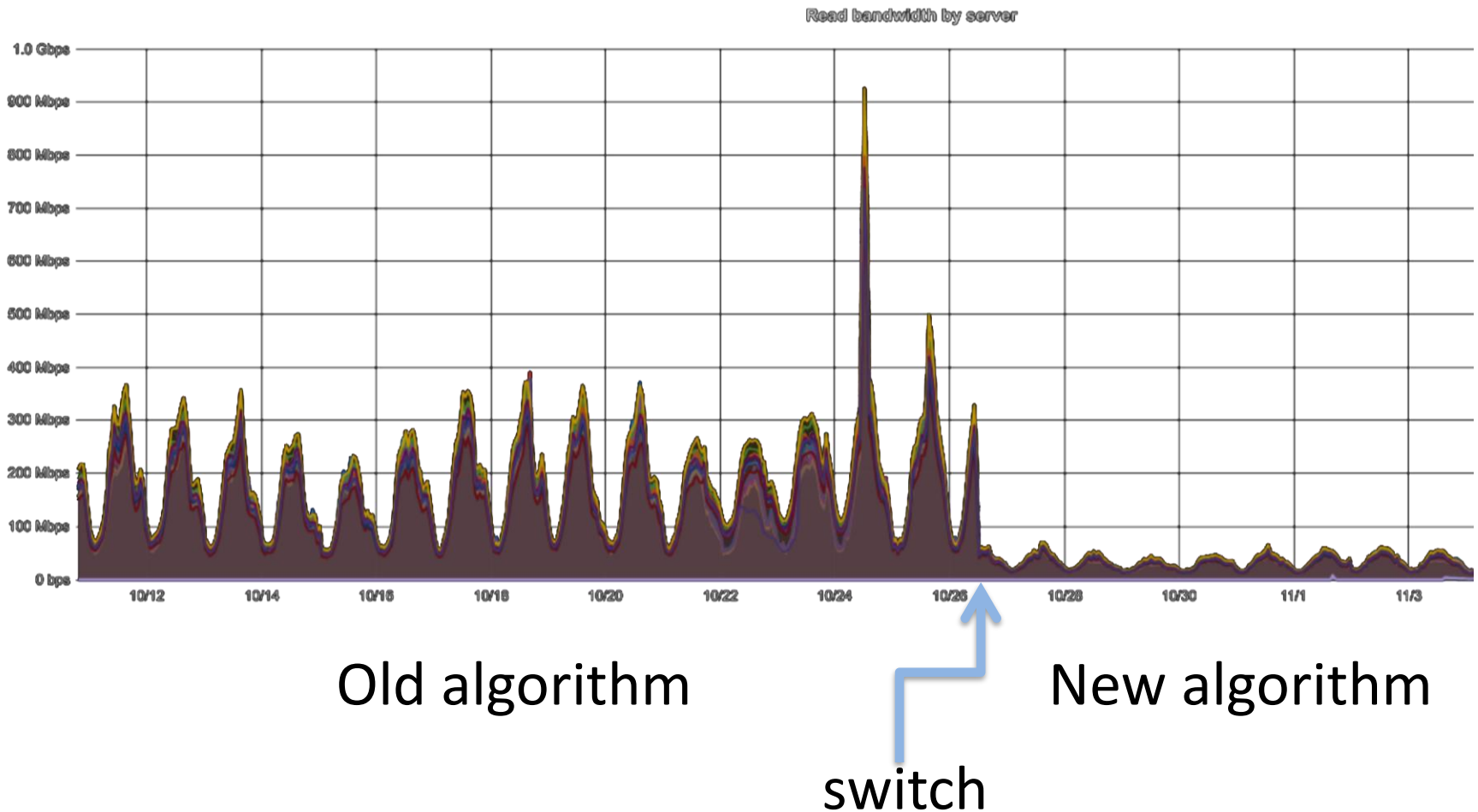
Computer Science > Data Structures and Algorithms

Consistent Hashing with Bounded Loads

Vahab Mirrokni, Mikkel Thorup, Morteza Zadimoghaddam

(Submitted on 3 Aug 2016)

Eliminating the bandwidth bottleneck



One algorithm, many applications

- Our algorithm has no details specific to video streaming. Works for *any* dynamic allocation system in the world – e.g. Google’s cloud.
- Mathematical analysis based on properties of degree-4 polynomials with random coefficients – the theory of which was originally developed with other applications in mind.

Lemma 10. *The expected number of balls hashing directly to any expected number of balls forwarded into q from its predecessor q^- is not active, and its active successor q^+ is given an extra capacity of one bin starting from q^+ is $O((\log c)/c^2)$.*

Proof. For the first statement, we note that the expected number of balls n/r for any $0 \leq i \leq r$. These are not added to q if some bin hash to $[h(q) - i, h(q) + i]$ event because balls and bins hash independently. The expected number of bins in $[h(q) - i, h(q) + i]$ is $\mu = i(n-1)/r$. For $i \geq r/(n-1)$, we have $\mu \geq 1$, and then, by the Chernoff bound, the probability of having a bin in $[h(q) - i, h(q) + i]$ is $O((\mu + \mu^2)/(\mu - 0)^4) = O(1/\mu^2) = O((r/(n-1))^2)$. The expected number of balls hashing directly to q is thus bounded by

$$n/r \cdot \left(\lfloor r/(n-1) \rfloor + \sum_{i=\lfloor r/(n-1) \rfloor + 1}^{\infty} (r/(ni))^2 \right)$$

We also have to consider the probability that the preceding bin q^- for q would need q^- to be filled even if we increased its capacity by 1 at least 2. This is bounded by the probability of having an interval $I \ni [h(q^-), h(q)]$ bins including one with capacity at least 2. This is what we analyzed in the previous section. $\Pr[d \geq 1] \leq \mathbf{E}[d] = O((\log c/c^2))$. By the capacity constraint, the number of balls forwarded to and end in q is $2cm/n$, so the expected number is

$$O((\log c/c^2)2cm/n) = O((m/n)(\log c/c^2))$$

Next we ask for the expected number d of full bins starting from the bin q , when q^+ is given an extra capacity of one. Again this implies that the analysis from the proof of Lemma 9 implies that $\mathbf{E}[d] = O((\log c/c^2))$.

Press on recent work with K.G. Larsen (Århus), J. Nelson (Harvard), Nguen (Northeastern)

The image shows a screenshot of the Wired website homepage. The browser address bar displays <https://www.wired.com>. The navigation menu includes categories: BUSINESS, CULTURE, DESIGN, GEAR, SCIENCE, SECURITY, and TRANSPORTATION. The main content area features several article cards. A blue circle highlights the article titled "Best-Ever Algorithm Found for Huge Streams of Data" by Kevin Hartnett, which is categorized under "BIG DATA". The article's image shows a stylized green figure on a red grid. Other visible articles include "Disaster-in-Waiting: After the M... Top Ash" by Adam Rogers, "Trump Released the JFK Files! Well, Sorta" by Graeme McMillan, and "How the Kodi Box Took Over Piracy" by Brian Barrett. A "MOST POPULAR" section on the right lists several other articles.

Wired

BUSINESS CULTURE DESIGN GEAR SCIENCE SECURITY TRANSPORTATION

WILDFIRES
Disaster-in-Waiting:
After the M...
Top Ash
ADAM ROGERS

WHILE YOU WERE OFFLINE
Trump Released the JFK Files! Well, Sorta
GRAEME MCMILLAN

BIG DATA
Best-Ever Algorithm Found for Huge Streams of Data
KEVIN HARTNETT

AUTONOMOUS VEHICLES
PREPPING ROBOCARS FOR THE WORLD'S MOST CHAOTIC CITIES
KAVEH WADDELL

MOST POPULAR

BUSINESS
Solve These Tough Data Problems and Watch Job Offers Roll In
TOM SIMONITE

SECURITY
How the Kodi Box Took Over Piracy
BRIAN BARRETT

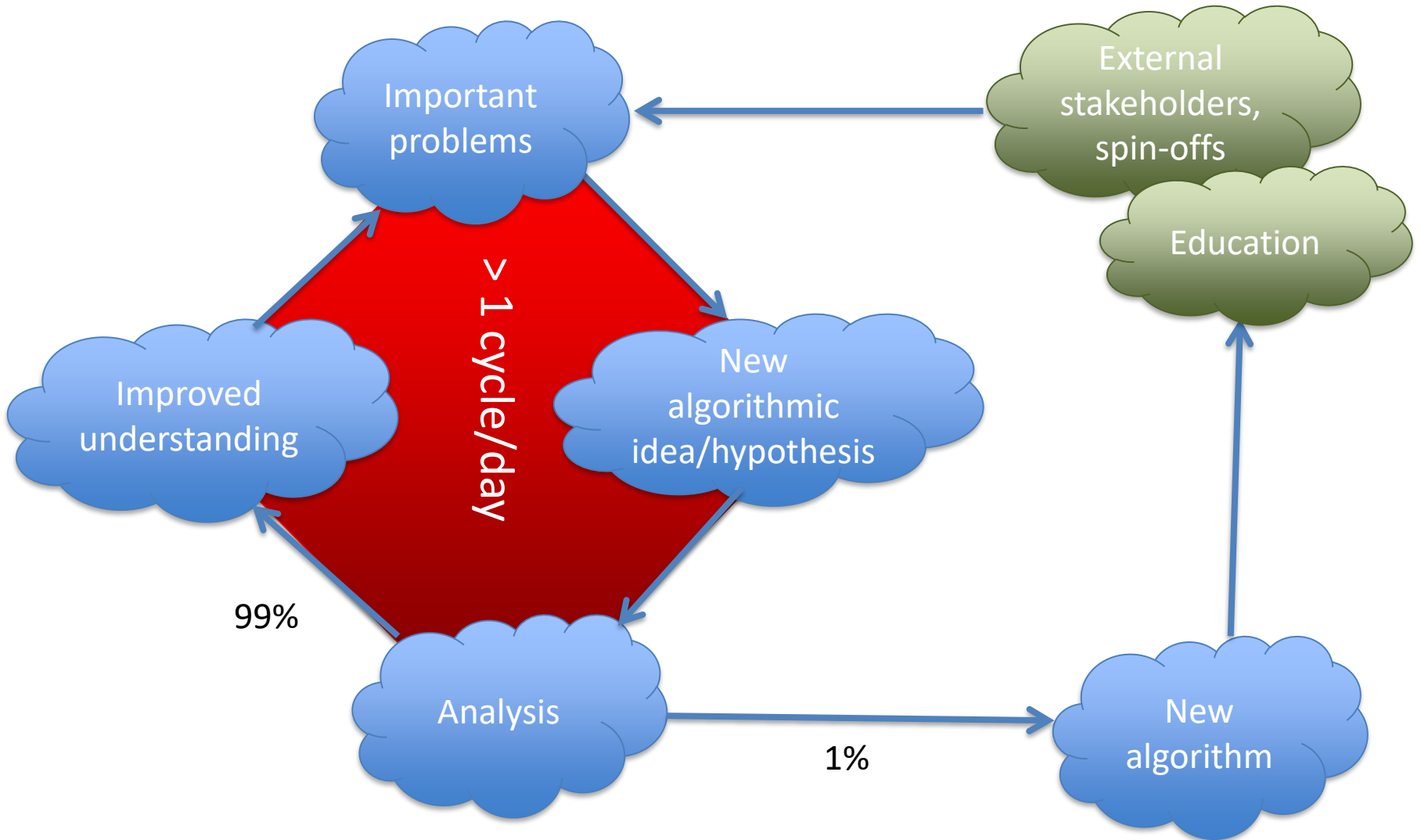
BUSINESS
Meet the High Schooler Shaking Up Artificial Intelligence
TOM SIMONITE

SCIENCE
Best-Ever Streaming Algorithm Found for Huge Amounts of Data
KEVIN HARTNETT

SCIENCE
Newfound Wormhole Allows Information to Escape Black Holes

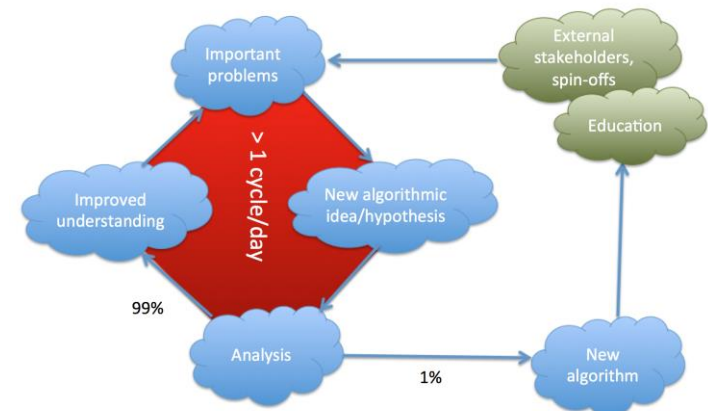
Nature of algorithmic research

- All research can be thought of as feedback loop,
 - Generating hypothesis/idea.
 - Experiment testing hypothesis, with outcome affecting next hypothesis.
- In experimental fields, experiments may be expensive and take years, so they better be carefully thought out.
- In algorithms we use mathematical thought experiments that may take a day or less...
This changes everything!



Ingredients of successful algorithmic research

- Talent.
- Deep understanding and intuition.
- Meeting of great minds in a creative environment.
- Inspiring leadership, research activities.
- Courage to do unpredictable research.



Core team



Thorup



Husfeldt



Pagh



Alstrup

Full professors with proven track record:

- **Academic impact** (H-indexes 54, 16, 26, 21)
- **Visibility** (keynotes ICDT '15, CPM '15, ESA '15, ECML '16, HALG '16, ICALP '17,...; editorial board/PC memberships of leading journals/conf.; “digitale vismandsråd” member...; appeared in BBC, ZDF, DR)
- **Managing large research projects** (DFF Adv. Grant 2013-18, ERC Consolidator 2014-19, Inno+ Big Data 2016-20)
- **Industrial impact** (AT&T Fellow Honor; Science Innovation Prize; founding Octoshape w/100M users,...)
- **Nurturing talent** (our PhD students received Presburger award, Nerode Prize, EliteForsk stipends)



Unique strength to attract the best talent

Vision

Ideal research environment,
attractive to the best
researchers in the world.

Regarded as the
leading center for
algorithms in Europe

BARC!

World-wide
impact on
science and
industry

High impact on
education

Effect multiplied by synergy
with Danish algorithms
community and industry

Flying start from DFF center Efficient Algorithms and Data Structures (EADS)

- Leadership in Algorithms
- ACM-SIAM Symposium on Discrete Algorithm (SODA) top international venue.
- 10 out of 180 papers accepted worldwide for SODA'18 from DIKU.
- I am 1 out of 4 invited plenary speakers.

PhD students

- 4 PhD students finished this year.
- Eva Rotenberg become Assistant Professor at DTU.
- Mikkel Abrahamsen joined 3Shape researching hearing aids.
- Søren Dahlgaard and Mathias Bæk Knudsen and doing machine learning start-up SupWiz.

Summary

- ICT is a vehicle for societal progress
- Designing efficient algorithms for fundamental problems is key to keeping up progress
- Algorithms research is best performed in a vibrant center where the best minds interact
- Unique strength of PI and core team can attract the best PhD students, post-docs, and visitors
- BARC aims to become the leading hub for algorithms research in Europe

